The Annals of Statistics
1973, Vol. 1, No. 6, 1105-1125

RANK TESTS FOR ONE SAMPLE, TWO SAMPLES, AND k SAMPLES WITHOUT THE ASSUMPTION OF A CONTINUOUS DISTRIBUTION FUNCTION[1]

BY W. J. CONOVER

Texas Tech University

The theory of rank tests has been developed primarily for continuous random variables. Recently the asymptotic theory of linear rank tests has been extended to include purely discrete random variables under the null hypothesis of randomness (including the two-sample and k-sample problems) and under contiguous alternatives, for the two methods of assigning scores known as the average scores method and the randomized ranks method.

In this paper the theory of rank tests is developed with no assumptions concerning the continuous or discrete nature of the underlying distribution function. Conditional rank tests, given the vector of ties, are shown to be similar, and the locally most powerful conditional rank test is given. The asymptotic distribution of linear rank statistics is given under the null hypotheses of randomness and symmetry (which includes the one-sample problem), and under contiguous alternatives. Three methods of assigning scores, the average scores, midranks, and randomized ranks methods, are discussed and briefly compared.

1. Introduction. Rank tests for randomness include such popular tests as the Wilcoxon-Mann-Whitney test, the Kruskal-Wallis test, the Fisher-Yates-Terry-Hoeffding (normal scores) test, and the van der Waerden (X) test. Rank tests for symmetry include the sign test and the Wilcoxon matched pairs signed ranks test, among others. Because of the difficulty in working with ranks in the presence of tied observations, most of the theoretical research involving rank tests begins with the assumption that all distribution functions are continuous. In applied work, however, ties do occur, and experimental research workers are sometimes hesitant to apply rank tests in the presence of many ties, for fear that the rank tests are no longer valid. The behavior of some rank tests, especially the Mann-Whitney test, in the presence of ties has been investigated by Chanda (1963), Putter (1955), Bühler (1967) and others.

The appearance of papers by Chernoff and Savage (1958) and by Hájek (1961, 1962) were significant in the development of the theory of rank tests, although the theory was restricted to continuous random variables, primarily under location and scale alternatives. More general alternatives were subsequently considered by Beran (1970), as well as by Chanda (1963), Andrews and Truax (1964),

Received June 1971; revised February 1973.

[1] Research sponsored by a Career Development Award, No. K4-GM-42, 351 from the Institute of General Medical Sciences, N.I.H. This research was conducted in part while the author was a visiting professor at the Seminar for Applied Mathematics, University of Zürich, and in part at Kansas State University.

Key words and phrases. Rank tests, ties, asymptotic efficiency.

and Adichie (1967), but still only for continuous random variables. In a recent paper by Vorlíčková (1970) the assumption of continuous random variables is replaced by the assumption that the random variables are purely discrete, and the asymptotic theory of linear rank statistics for testing randomness is developed along the lines of Hájek and Šidák (1967).

In this paper the theory of rank tests for randomness and rank tests for symmetry is developed along the lines of Hájek and Šidák (1967), except that the assumption of continuous distribution functions is never made. Several popular methods of handling ties are examined, and some conditions are stated under which large sample approximate normality holds for linear rank statistics.

In Section 2, the conditional distribution of the vector of ranks is given under the null hypothesis H_0 that all observations are from independent and identically distributed random variables, given the vector of ties. This conditional distribution is shown to be independent of the population distribution function. A similar development is given in Section 3 for the conditional distribution of ranks under a different null hypothesis H_1 which states that the observations come from independent and identically distributed random variables whose distribution is symmetric about some point.

Then linear rank statistics are introduced. Conditions which are sufficient for asymptotic normality of linear rank statistics are given in Sections 4 and 5 for H_0 and H_1 respectively. These conditions are met when ties among the observations are handled using midranks, average scores, or randomized ranks under some more easily verified conditions stated in Theorems 4.2, 4.3, and 4.4.

Alternative hypotheses are considered in subsequent sections. The alternative to H_0 considered here is nonidentical distributions, such as in the usual two sample and k sample problems. The alternative to H_1 is nonsymmetry. The locally most powerful (LMP) rank tests for H_0 and H_1 are shown to be based on linear rank statistics, in Sections 6 and 7.

In Sections 8 and 9 the asymptotic distributions are derived for the LMP test statistics given in Sections 6 and 7, and for a large class of other linear rank statistics. Asymptotic normality is shown under a set of assumptions which assures that the sequence of alternatives is contiguous to the sequence of null hypotheses. Asymptotic relative efficiency for linear rank tests is discussed in Sections 10 and 11, with particular attention paid to the three methods of handling ties discussed in Section 4. Some applications are briefly discussed in Section 11.

2. The exact distribution of rank statistics under H_0. The *rank* R_i of a random variable X_i is the number of X_j's less than or equal to X_i in a realization $\mathbf{x} = (x_1, \cdots, x_N)$ of the vector random variable $\mathbf{X} = (X_1, \cdots, X_N)$. The ith *order statistic* $X^{(i)}$ is the random variable which assumes the value $x^{(i)}$ in the ordered realization $\mathbf{x}^{(\cdot)} = (x^{(1)} \leqq \cdots \leqq x^{(N)})$ of $\mathbf{x}$. Then the rank of $X^{(i)}$ is an integer greater than or equal to i.

We say $x^{(j+1)}$ is in a *tie of length* τ_i if $x^{(j)} < x^{(j+1)} = x^{(j+2)} = \cdots = x^{(j+\tau_i)} < x^{(j+\tau_i+1)}$, for $\tau_i \geqq 1$. Let $\boldsymbol{\tau}(\mathbf{x}) = (\tau_1, \cdots, \tau_k)$ denote the sequence of lengths of ties in $\mathbf{x}^{(\cdot)}$. The vector random variable $\mathbf{R} = (R_1, \cdots, R_N)$ depends on the vector $\boldsymbol{\tau}(\mathbf{x})$ in that for a given $\mathbf{x}$ the ranks assume only the values $t_i = \sum_{j=1}^{i} \tau_j$, $i = 1, \cdots, k$. That is, for a given $\boldsymbol{\tau}(\mathbf{x})$, $\mathbf{R}$ is restricted to some permutation of the vector $(t_1, \cdots, t_k)$ where each element t_i appears τ_i times. Let $\{\mathbf{R} \mid \boldsymbol{\tau}(\mathbf{x})\}$ be the space of all distinct permutations of $\mathbf{R}$ given $\boldsymbol{\tau}(\mathbf{x})$.

By H_* we denote the hypothesis that components of $\mathbf{X}$ are interchangeable: $H_*: P(X_1 \leqq x_1, \cdots, X_N \leqq x_N) = P(X_1 \leqq x_{i_1}, \cdots, X_N \leqq x_{i_N})$ holds for each permutation $(i_1, \cdots, i_N)$ of $(1, \cdots, N)$. This is more general than the hypothesis of randomness, which is denoted by $H_0: P(X_1 \leqq x_1, \cdots, X_N \leqq x_N) = \prod_{i=1}^{N} P(X_1 \leqq x_i)$.

The following theorem is analogous to II.1.2.a of [10]. The proof is combinatoric in nature and is omitted.

THEOREM 2.1. *Let H_* be true. Then the distribution function $\bar{F}(\mathbf{x}^{(\cdot)})$ of $\mathbf{X}^{(\cdot)} = (X^{(1)} \leqq \cdots \leqq X^{(N)})$ is absolutely continuous with respect to the distribution function $F(\mathbf{x})$ of $\mathbf{X}$. The Radon–Nikodym derivative of $\bar{F}(\mathbf{x}^{(\cdot)})$ with respect to $F(\mathbf{x})$ exists and satisfies*

$$\frac{d\bar{F}(\mathbf{x}^{(\cdot)})}{dF(\mathbf{x}^{(\cdot)})} = \binom{N}{\tau}, \qquad \text{where} \quad \binom{N}{\tau} = \frac{N!}{\tau_1! \cdots \tau_k!}. \tag{2.1}$$

Also

$$\begin{aligned} P(\mathbf{R} = \mathbf{r} \mid \boldsymbol{\tau}(\mathbf{x})) &= 1/\binom{N}{\tau} && \text{if} \quad \mathbf{r} \in \{\mathbf{R} \mid \boldsymbol{\tau}(\mathbf{x})\} \\ &= 0 && \text{if} \quad \mathbf{r} \notin \{\mathbf{R} \mid \boldsymbol{\tau}(\mathbf{x})\} \end{aligned} \tag{2.2}$$

and for a given $\boldsymbol{\tau}(\mathbf{x})$, $\mathbf{R}$ is independent of $\mathbf{X}^{(\cdot)}$.

Theorem 2.1 states that all conditional rank tests, given the vector of ties, are similar over H_*, and hence are distribution free. The assumption of a continuous distribution function has the effect of furnishing the vector of ties, namely $(\tau_1, \cdots, \tau_k) = (1, \cdots, 1)$ with probability 1. Rank tests without the vector of ties given are in general not similar for H_* or for H_0, a fact which is easy to show by trivial examples.

3. The exact distribution of rank statistics under the null hypothesis of symmetry H_1. The *rank* R_i^+ of the absolute value of a random variable X_i is the number of $|X_j|$'s less than or equal to $|X_i|$ in a realization $|\mathbf{x}| = (|x_1|, \cdots, |x_N|)$ of the vector random variable $|\mathbf{X}| = (|X_1|, \cdots, |X_N|)$. The ith *order statistic* $|X|^{(i)}$ is the random variable which assumes the value $|x|^{(i)}$ in the ordered realization $|\mathbf{x}|^{(\cdot)} = (|x|^{(1)} \leqq \cdots \leqq |x|^{(N)})$ of $|\mathbf{X}|^{(\cdot)}$. Also, $|x|^{(j+1)}$ is in a tie of length τ_i if $|x|^{(j)} < |x|^{(j+1)} = |x|^{(j+2)} = \cdots = |x|^{(j+\tau_i)} < |x|^{(j+\tau_i+1)}$, for $\tau_i \geqq 1$. An exception arises because we let τ_0 denote the number of observations equal to zero, so that τ_0 may equal zero. Let $\boldsymbol{\tau}(|\mathbf{x}|) = (\tau_0, \cdots, \tau_k)$ denote the sequence of lengths of ties in $|\mathbf{x}|^{(\cdot)}$. The vector $\mathbf{R}^+ = (R_1^+, \cdots, R_N^+)$ depends on the vector $\boldsymbol{\tau}(|\mathbf{x}|)$ in that for a given $|\mathbf{x}|$ the ranks assume only the values $t_i = \sum_{j=0}^{i} \tau_j$, $i = 1, \cdots, k$, and $t_0 = \tau_0$ if $\tau_0 > 0$. That is, for a given $\boldsymbol{\tau}(|\mathbf{x}|)$, R^+ is restricted to some permutation of the vector $(t_0, \cdots, t_k)$ where each element t_i appears τ_i times. A realization of $\mathbf{R}^+$ is denoted by $\mathbf{r}^+ = (r_1^+, \cdots, r_N^+)$. Let $\{\mathbf{R}^+ \mid \boldsymbol{\tau}(|\mathbf{x}|)\}$ be the space of all distinct permutations of $\mathbf{R}^+$ given $\boldsymbol{\tau}(|\mathbf{x}|)$.

Consider the function sign $x = 1$, 0 or -1 depending on whether x is positive, zero, or negative, and introduce the sign statistics

$$\text{sign } \mathbf{X} = (\text{sign } X_1, \cdots, \text{sign } X_N) .$$

Then X_i equals sign $X_i \cdot |X|^{(R_i^+)}$. Also, let $V(\mathbf{r}^+, \tau_0)$ be the space of N dimensional vectors $\mathbf{v}$ whose elements v_i are 0 if $r_i^+ = \tau_0$, and are either $+1$ or -1 if $r_i^+ > \tau_0$, where $\mathbf{r}^+$ and τ_0 are obtained from the same vector of ties. Note that for every $\mathbf{r}^+$ there is a space V with $2^{N-\tau_0}$ distinct vectors.

By H_1 we denote the null hypothesis of symmetry. That is, H_1 states that $X_1, \cdots, X_N$ are i.i.d. according to a distribution function $F(x)$ which is symmetric about zero in the sense that $P(X_i < x) = P(X_i > -x)$ for all x. The assumption of symmetry may be stated as the Radon–Nikodym derivative of $F(-x)$ with respect to $F(x)$ exists and equals -1 almost everywhere with respect to $F(x)$. Further denote the distribution functions of $|X_i|$, $\mathbf{X}$, $|\mathbf{X}|$, and $|\mathbf{X}|^{(\cdot)}$ by $F^+(x)$, $F(\mathbf{x})$, $F^+(\mathbf{x})$, and $\bar{F}^+(\mathbf{x})$ respectively. The following theorem is analogous to Theorem II.1.3 of [10]. The proof is omitted.

THEOREM 3.1. *Let H_1 be true. Then $\bar{F}^+(|\mathbf{x}|^{(\cdot)})$ is absolutely continuous with respect to $F(|\mathbf{x}|)$ and the Radon–Nikodym derivative satisfies*

$$\frac{d\bar{F}^+(|\mathbf{x}|^{(\cdot)})}{dF(|\mathbf{x}|)} = 2^{N-\tau_0}\binom{N}{\tau} , \qquad \textit{where} \quad \binom{N}{\tau} = \frac{N!}{\tau_0! \cdots \tau_k!} , \tag{3.1}$$

and

$$P(\mathbf{R}^+ = \mathbf{r}^+, \text{sign } \mathbf{X} = \mathbf{v} \mid \boldsymbol{\tau}(|\mathbf{x}|)) = (\tfrac{1}{2})^{N-\tau_0}/\binom{N}{\tau} \tag{3.2}$$

if

$$\mathbf{r}^+ \in \{\mathbf{R}^+ \mid \boldsymbol{\tau}(|\mathbf{x}|)\} \quad \textit{and} \quad \mathbf{v} \in V(\mathbf{r}^+, \tau_0) , \qquad = 0 \quad \textit{elsewhere} .$$

The principal conclusion that may be drawn from Theorem 3.1 is that all conditional rank tests based on $\mathbf{R}^+$ and sign $\mathbf{X}$, given the vector of ties $\boldsymbol{\tau}(|\mathbf{x}|)$, are similar over H_1 and hence are distribution free. This follows immediately from (3.2). The assumption of a continuous distribution function has the effect of furnishing the vector of ties, namely $\tau_0 = 0$ and $\tau_i = 1$ for $1 \leq i \leq N$, with probability 1. If one prefers to choose a test based on the ranks R_i^+ of only those observations where sign $X_i = 1$, then the following version of Theorem IV.1.4.a. of [10] may be of interest.

THEOREM 3.2. *Let H_1 be true, and let I be any subset of the set $\{t_1, \cdots, t_N\}$ where each t_i appears τ_i times. The probability that the set of ranks R_i^+ of positive observations X_i, given $\tau(|\mathbf{x}|)$, will coincide with the set I (without regard to order) is equal to $(\frac{1}{2})^{N-\tau_0}$.*

PROOF. The set $\{t_1, \cdots, t_N\}$ is the set of ranks belonging to the nonzero $|X_i|$'s. Since the signs of the nonzero X_i's are independent of their absolute values, the result follows easily.

4. The asymptotic distribution of linear rank statistics under H_0. Our attention will be restricted to a useful class of statistics called linear rank statistics. Linear rank statistics, where ties are possible, are defined by

$$S_c = \sum_{i=1}^N c_i a_N(R_i; T_N) \tag{4.1}$$

where $a_N(R_i; T_N)$ are scores dependent only on N, R_i, and the *rank empirical distribution function* $T_N(u)$ defined by

$$T_N(u) = \frac{1}{N}\{\text{number of } R_i\text{'s} \leqq uN\}, \qquad u \in [0, 1]. \tag{4.2}$$

In the case of no ties, $N \cdot T_N(u)$ simply equals $[uN]$, using the "greatest integer" notation. The constants c_i are called regression constants. It is easy to show from (2.2) that under H_* we have

$$E\{S_c \mid T_N\} = \frac{1}{N} \sum_{i=1}^N c_i \sum_{j=1}^N a_N(r_j; T_N) = N\bar{c}\bar{a}, \tag{4.3}$$

and

$$\operatorname{Var}\{S_c \mid T_N\} = \frac{1}{N-1} \sum_{i=1}^N (c_i - \bar{c})^2 \sum_{j=1}^N (a_N(r_j; T_N) - \bar{a})^2. \tag{4.4}$$

For all functions we will adopt the convention,

$$f^{-1}(t) = \inf\{x \mid f(x) \geqq t\}. \tag{4.5}$$

The theorems of this section present conditions under which

$$\frac{S_c - E\{S_c \mid T_N\}}{[\operatorname{Var}\{S_c \mid T_N\}]^{\frac{1}{2}}} \to N(0, 1) \tag{4.6}$$

holds, where $\to N(0, 1)$ means "has asymptotically the standard normal distribution function." Throughout, we let $\phi(u)$ denote an arbitrary function defined on the interval $0 \leqq u \leqq 1$. Some of the conditions common to the theorems in this section are as follows:

$$H_0 \text{ is true}; \tag{4.7}$$

$$0 < \int_0^1 (\phi(u) - \bar{\phi})^2\, du < \infty, \qquad \text{where } \bar{\phi} = \int_0^1 \phi(u)\, du; \tag{4.8}$$

$$\sum_{i=1}^N (c_i - \bar{c})^2 / \max_{1 \leqq i \leqq N} (c_i - \bar{c})^2 \to \infty. \tag{4.9}$$

The following theorem is a generalization of V.1.6.a. of [10].

THEOREM 4.1. *If conditions* (4.7), (4.8), (4.9), *and*

$$\int_0^1 (a_N(N \cdot T_N^{-1}(u); T_N) - \phi(u))^2\, du \to_P 0 \tag{4.10}$$

hold, then (4.6) *follows.*

REMARK. Condition (4.10) is not easy to verify, but the major difficulty is removed with the aid of Theorems 4.2, 4.3 and 4.4 which follow the proof of this theorem.

PROOF. Consider the random variables $Y_i = F(X_i)$ which under (4.7) are i.i.d. with some cdf $G(u)$. Let $W_1, \cdots, W_N$ be i.i.d. uniform random variables which are also independent of the Y_i. Let $G(\{\cdot\})$ be the measure induced by $G(u)$ on any set $\{\cdot\}$ of real numbers. Then $G(\{y\})$ equals $P(Y = y)$ at discontinuity points of $G(u)$, and equals zero elsewhere. Thus the random variables

$$U_i = Y_i - W_i G(\{Y_i\})$$

are mutually independent with the uniform distribution on (0, 1), according to the following reasoning. Let

$$a(u) = G(G^{-1}(u)) - G(\{G^{-1}(u)\})$$

and

$$b(u) = G(G^{-1}(u)) .$$

Then

$$P(U_i \leqq u) = P(Y_i \leqq b(u),\ W_i G(\{G^{-1}(u)\}) \geqq b(u) - u) .$$

If $G(u) = u$, then $b(u) = u$ and $P(U_i \leqq u) = P(Y_i \leqq b(u)) = u$. If $G(u) < u$, then $G(u)$ is constant on the interval $[a(u), b(u))$ and $W_i G(\{G^{-1}(u)\})$ is uniformly distributed on $(0, b(u) - a(u))$. This leads to

$$\text{(4.11)} \quad \begin{aligned} P(U_i \leqq u) &= P(U_i \leqq a(u)) + P(a(u) < U_i \leqq u) \\ &= a(u) + P(Y_i = b(u))P\left(W_i \geqq \frac{b(u) - u}{b(u) - a(u)}\right) = u . \end{aligned}$$

It is shown in [10], page 153, that under assumptions (4.8) and (4.9) the random variable T_c/σ_c, where

$$\text{(4.12)} \quad T_c = \textstyle\sum_{i=1}^N (c_i - \bar{c})\phi(U_i) ,$$

and

$$\text{(4.13)} \quad \sigma_c^2 = \textstyle\sum_{i=1}^N (c_i - \bar{c})^2 \int_0^1 (\phi(u) - \bar{\phi})^2 \, du$$

has asymptotically the standard normal distribution function. It is also shown in [10] page 160 that S_c^ϕ/σ_c, where

$$\text{(4.14)} \quad S_c^\phi = \textstyle\sum_{i=1}^N (c_i - \bar{c}) a_N^\phi(R_i^*) ; \qquad a_N^\phi(i) = E\{\phi(U_1) \mid R_1^* = i\} ,$$

$$\text{(4.15)} \quad R_i^* = \text{the rank of } U_i ,$$

satisfies

$$\text{(4.16)} \quad E\left\{\frac{(S_c^\phi - T_c)^2}{\sigma_c^2}\right\} \to 0 ,$$

and hence $S_c^\phi/\sigma_c \to N(0, 1)$ under (4.8) and (4.9).

We have, by simple algebra and by (4.4),

$$
\begin{aligned}
& E\{[S_c - E\{S_c \mid T_N\} - S_c{}^\phi]^2 \mid T_N\} \\
& \quad = E\{[\textstyle\sum_{i=1}^N (c_i - \bar{c})(a_N(R_i; T_N) - a_N{}^\phi(R_i{}^*))]^2 \mid T_N\} \\
(4.17) \qquad & \quad \leqq \frac{1}{N-1} \textstyle\sum_{i=1}^N (c_i - \bar{c})^2 \sum_{j=1}^N [a_N(r_j; T_N) - a_N{}^\phi(r_j{}^*)]^2 \\
& \quad = \frac{N}{N-1} \textstyle\sum_{i=1}^N (c_i - \bar{c})^2 \int_0^1 [a_N(N \cdot T_N{}^{-1}(u); T_N) \\
& \qquad\qquad - a_N{}^\phi(1 + [uN])]^2 \, du \,.
\end{aligned}
$$

Now

$$
(4.18) \qquad E\left\{\frac{[S_c - E\{S_c \mid T_N\} - S_c{}^\phi]^2}{\sigma_c{}^2}\right\} \to_P 0
$$

holds if the integral in (4.17) converges to zero in probability. But due to an algebraic inequality the integral in (4.17) is less than or equal to the sum

$$
(4.19) \qquad 2 \int_0^1 [a_N(N \cdot T_N{}^{-1}(u); T_N) - \phi(u)]^2 \, du + 2 \int_0^1 [a_N{}^\phi(1 + [uN]) - \phi(u)]^2 \, du \,.
$$

The first integral in (4.19) converges to zero in probability by assumption, and the other is shown to converge to zero in [10] page 158. The rest of the proof follows the same lines as in [10], page 161.

In the usual hypothesis test using ranks a set of scores $a_N(i)$, $i = 1, \cdots, N$, is suggested, where the scores are "well behaved" in some sense. The scores are then assigned to the observations X_j on the basis of their ranks R_j. The assignment of scores to the observations is unique if there are no ties. But if ties are present scores may be assigned to the observations in many different ways. Some of the more popular methods are known as the average score method, the midrank method, and the method of randomized ranks. The remainder of this section examines the connection between these methods and Theorem 4.1.

Average score method. We can assign to each random variable in a tie the average of the scores that belong to the random variables in the tie:

$$
(4.20) \qquad a_N(t_i; T_N) = \frac{1}{\tau_i} \textstyle\sum_{j=1}^{\tau_i} a_N(t_{i-1} + j) \,.
$$

For such a method the null distribution of S_c has been given in [14] for the case of purely discrete distributions. The general case is given in Theorem 4.2.

Let $\phi_a(u)$ be $\phi(u)$ averaged over the intervals in which $G(u)$ is constant valued:

$$
\begin{aligned}
(4.21) \qquad \phi_a(u) &= \phi(u) \qquad \text{if} \quad G(\{G^{-1}(u)\}) = 0 \,, \\
&= \frac{1}{b(u) - a(u)} \int_{a(u)}^{b(u)} \phi(t) \, dt \quad \text{otherwise} \,,
\end{aligned}
$$

where $a(u) = G^{-1}(u) - G(\{G^{-1}(u)\})$ and $b(u) = G^{-1}(u)$ are the left and right endpoints of the interval containing u.

THEOREM 4.2. *If conditions* (4.7), *and* (4.9) *hold, if the scores* $a_N(i)$ *satisfy*

$$\int_0^1 (a_N(1 + [uN]) - \phi(u))^2 \, du \to 0 \tag{4.22}$$

and if $\phi_a(u)$ *is square integrable and non constant over* (0, 1), *then* (4.6) *holds for average scores defined by* (4.20).

PROOF. Because of Theorem 4.1 it is only necessary to show that (4.22) implies (4.10), which takes the following form

$$\int_0^1 (a_N(N \cdot T_N^{-1}(u); T_N) - \phi_a(u))^2 \, du \to_P 0 \tag{4.23}$$

for scores defined by (4.20).

Let $\phi_T(u)$ be $\phi(u)$ averaged over the intervals in which $T_N(u)$ is constant valued:

$$\phi_T(u) = \frac{1}{d(u) - c(u)} \int_{c(u)}^{d(u)} \phi(t) \, dt \tag{4.24}$$

where $c(u) = T_N^{-1}(u) - T_N(\{T_N^{-1}(u)\})$ and $d(u) = T_N^{-1}(u)$ are the left and right endpoints of the interval containing u. Then $a(N \cdot T_N^{-1}(u); T_N) - \phi_T(u)$ is merely $a(1 + [uN]) - \phi(u))$ averaged over the intervals of constant $T_N(u)$. For each of these intervals

$$\int (a(N \cdot T_N^{-1}(u); T_N) - \phi_T(u))^2 \, du \leqq \int (a(1 + [uN]) - \phi(u))^2 \, du \tag{4.25}$$

holds, and hence (4.25) holds over the entire unit interval. Both integrals in (4.25) converge to zero by assumption (4.22). Since the integral in (4.23) is $\leqq$ the sum

$$2 \int_0^1 (a(N \cdot T_N^{-1}(u); T_N) - \phi_T(u))^2 \, du + 2 \int_0^1 (\phi_a(u) - \phi_T(u))^2 \, du \tag{4.26}$$

it remains to show that the latter integral converges to zero in probability.

The proof that the latter integral converges to zero in probability is long and not very elegant. Basically, the integral over the intervals where $\phi_a(u) = \phi(u)$ converges to zero by Lemma V.1.6.b. of [10]. The integral over the remaining intervals converges to zero in probability for the same reasons the integral B_N on page 280 of [14] converges to zero in probability. The reader is spared the details.

Note that because of Lemma V.1.6.a. of [10] we may let $a_N(i)$ be defined by

$$a_N(i) = \phi\left(\frac{i}{N + 1}\right) \tag{4.27}$$

if $\phi(u)$ is expressible as a finite sum of square integrable and monotone functions, and (4.22) then holds.

Midrank method. If the scores $a_N(i)$ are defined for half integer i as well as integer i then it may be more convenient to use the scores corresponding to the average (mid) rank,

$$a_N(t_i; T_N) = a_N\left\{\frac{t_{i-1} + t_i + 1}{2}\right\}. \tag{4.28}$$

The following theorem gives some conditions under which S_c approaches normality. Let $\{I_k\}_{k\geqq 0}$ denote the denumerable set of discontinuity intervals $(a(u), b(u)]$, where $a(u)$ and $b(u)$ are defined as in (4.21) for each discontinuity point of $G(u)$. Let $\phi_m(u)$ equal $\phi(u)$ if u is in a continuity interval, and let $\phi_m(u) = \phi(\text{med } I_j)$ if u is in a discontinuity interval I_j, where med I_j refers to the midpoint of I_j, $(a(u) + b(u))/2$.

THEOREM 4.3. *Let conditions* (4.7), (4.9) *and* (4.22) *hold. If* $\phi_m(u)$ *is square integrable and non constant over* (0, 1), *if* $\{\text{med } I_k\}_{k\geqq 0}$ *are continuity points of* $\phi(u)$, *and if*

$$a_N\left(\frac{1+[2uN]}{2}\right) \to \phi(u) \qquad \textit{for } 0 < u < 1\,, \tag{4.29}$$

then (4.6) *holds for the midrank scores* (4.28).

PROOF. In view of Theorem 4.1, it suffices to show that (4.10), which takes the form

$$\textstyle\int_0^1 (a_N(N \cdot T_N^{-1}(u); T_N) - \phi_m(u))^2\, du \to_P 0\,, \tag{4.30}$$

holds for scores defined by (4.28).

Again, we present only an abbreviated form of the proof. The integral in (4.30), over the intervals in which $\phi_m(u) = \phi(u)$, converges to zero in probability primarily because of (4.22). The remaining portions of the integral converge to zero in probability because of (4.29) and because $\phi(u)$ is continuous at $u = (\text{med } I_k)_{k\geqq 0}$.

Note that Lemma V.1.6.a. of [10] is equally valid if the scores are defined by

$$a_N\left(\frac{k}{2}\right) = \phi\left(\frac{k/2}{N+1}\right); \qquad 2 \leqq k \leqq 2N \tag{4.31}$$

where $\phi(u)$ is expressible as a finite sum of square integrable and monotone functions, and hence (4.22) holds. These conditions do not eliminate the need for assuming that med I_j are continuity points of $\phi(u)$, and that $\phi_m(u)$ is non constant, but the square integrability assumption for $\phi_m(u)$, and (4.29), may now be inferred.

Randomized ranks. In case several Y's have the same rank t_ι, a one to one correspondence of the integers $\{t_{i-1} + j\}_{j=1}^{\tau_i}$ with the τ_i tied Y's may be established on the basis of some independent experiment which gives each possible assignment of the integers, then called *randomized ranks* R^*, equal probability. One such method of assigning randomized ranks is discussed in the proof of Theorem 4.1. Asymptotic normality of the resulting S_c is easily shown.

THEOREM 4.4. *If* (4.7), (4.8), (4.9) *and* (4.22) *hold, then* (4.6) *follows for scores given by*

$$a_N(R_i; T_N) = a_N(R_i^*) \tag{4.32}$$

where R_i^* *are randomized ranks.*

PROOF. Since $a_N(N \cdot T_N^{-1}(u); T_N) = a_N(1 + [uN])$, (4.22) implies (4.10), which because of Theorem 4.1 proves the theorem.

While the above theorems are sufficient for testing most two-sample problems, multi sample problems usually employ so-called Q statistics [10]. Let $\{s_1, \cdots, s_k\}$ be a partition of $(1, \cdots, N)$ and let $n_j = \text{card } s_j$. The Q statistic is defined as

$$Q_{n_1 \cdots n_k} = \frac{(N-1) \sum_{j=1}^{k} (S_{Nj} - n_j \bar{a})^2 / n_j}{\sum_{i=1}^{N} (a_N(R_i; T_N) - \bar{a})^2} \tag{4.33}$$

where

$$S_{Nj} = \sum_{i \in s_j} a_N(R_i; T_N) \tag{4.34}$$

and $\bar{a}$ is the average score as in (4.3). The following theorem is the noncontinuous analogue to Theorem V.2.2 of [10].

THEOREM 4.5. *Conditions* (4.7), (4.8), (4.10) *and* $\min(n_1, \cdots, n_k) \to \infty$ *imply* $Q_{n_1 \cdots n_k}$ *has asymptotically the chi-square distribution with* $k - 1$ *degrees of freedom.*

PROOF. The proof is like that in [10], except Theorem 4.1 is invoked instead of Theorem V.1.6.a. of [10].

The previous theorems of this section may be used to obtain scores which satisfy (4.10).

5. The asymptotic distribution of linear rank statistics under H_1. With H_1 we shall consider linear statistics of the following type,

$$S_N^+ = \sum_{i=1}^{N} a_N(R_i^+; T_N^+) \text{ sign } X_i \tag{5.1}$$

where $a_N(R_i^+; T_N^+)$ are scores dependent only on N, R_i^+, and the rank empirical distribution function $T_N^+(u)$ defined by

$$T_N^+(u) = \frac{1}{N} \{\text{number of } R_i^+\text{'s} \leqq uN\} \tag{5.2}$$

in a manner similar to that used in Section 4. Obviously under H_1 $E(S_N^+ \mid T_N^+) = 0$, $E(S_N^+) = 0$, and

$$\begin{aligned} \text{Var}(S_N^+ \mid T_N^+) &= \sum_{i=1}^{N} E\{[a_N(R_i^+; T_N^+) \text{ sign } X_i]^2 \mid T_N^+\} \\ &= \sum_{r_i^+ > \tau_0} a_N^2(r_i^+; T_N^+) . \end{aligned} \tag{5.3}$$

The following theorem is the noncontinuous counterpart to Theorem V.1.7. of [10].

THEOREM 5.1. *Let $\phi^+(u)$ be a square integrable (on $0 \leqq u \leqq 1$) function with*

$$\int_{F^+(0)}^{1} [\phi^+(u)]^2 \, du > 0 \tag{5.4}$$

and assume H_1 is true. Then satisfaction of

$$\int_{\tau_0/N}^{1} [a_N(N \cdot T_N^{+-1}(u); T_N^+) - \phi^+(u)]^2 \, du \to_P 0 \tag{5.5}$$

implies the sequence S_N^+/σ_N is asymptotically $N(0, 1)$ where

$$\sigma_N^2 = \sum_{r_i^+ > \tau_0} a_N^2(r_i^+; T_N^+) \cong N \int_{F^+(0)}^{1} [\phi^+(u)]^2 \, du = \sigma^2 . \tag{5.6}$$

PROOF. As in Theorem 4.1 put $Y_i^+ = F^+(|X_i|)$ where F^+ denotes the cdf of $|X_i|$. Let $U_i^+ = Y_i^+ - W_i G(\{Y_i^+\})$, where $G(u)$ is the cdf of Y_i^+, and $W_1, \cdots, W_N$ are independent random variables uniformly distributed on (0, 1).
Let

$$a_N^+(i) = E[\phi^+(U_1^+) \mid R_1^* = i] \tag{5.7}$$

and consider

$$S_\phi^+ = \sum_{i=1}^N a_N^+(R_i^*) \operatorname{sign} X_i \tag{5.8}$$

where R_i^* is the rank of U_i^+. Then $S_\phi^+/\sigma \to N(0, 1)$ as $N \to \infty$, as stated in Theorem 2 of [15].

Next we need to show $E\{S_N^+ - S_\phi^+)^2/\sigma^2\} \to_P 0$. We have

$$\begin{aligned} E\{(S_N^+ - S_\phi^+)^2 \mid T_N^+\} & \\ &= \operatorname{Var}(S_N^+ - S_\phi^+ \mid T_N^+) \\ &= \textstyle\sum_{r_i^+ > \tau_0} [a_N(r_i^+; T_N^+) - a_N^+(r_i^*)]^2 \\ &= N \textstyle\int_{\tau_0/N}^1 (a_N(N \cdot T_N^{+-1}(u); T_N^+) - a_N^+(1 + [uN]))^2 \, du \\ &\leqq 2N \textstyle\int_{\tau_0/N}^1 (a_N(N \cdot T_N^{+-1}(u); T_N^+) - \phi^+(u))^2 \, du \\ &\quad + 2N \textstyle\int_{\tau_0/N}^1 (a_N^+(1 + [uN]) - \phi^+(u))^2 \, du \,. \end{aligned} \tag{5.9}$$

The first integral in (5.9) converges in probability to zero by assumption (5.5) and the second converges to zero by Theorem V.1.4.b. of [10], page 158. Therefore

$$E\left\{\frac{(S_N^+ - S_\phi^+)^2}{\sigma^2}\right\} \to_P 0\,. \tag{5.10}$$

It remains to show that σ_N^2 converges stochastically to σ^2. But this is shown in the same manner as on page 161 of [10].

As with the tests for randomness, rank tests for symmetry are usually defined in terms of a set of scores $a(i)$, $i = 1, \cdots, N$. If ties exist in the absolute values of the data, there are several different ways of assigning scores to the data. The average scores method, the midrank method, or the randomized ranks method, as described in the previous section, may be used to assign scores to tied observations. For each method equation (5.5) holds, and theorems analogous to Theorems 4.2, 4.3, and 4.4 may be stated and proved as before. The details are obvious and are omitted in the interest of brevity. Results along this line for purely discrete parent distribution functions are given in [15].

6. The locally most powerful conditional rank test of H_0. The principal result of this section is an analogue to Theorem II.4.8. of [10] for possibly noncontinuous distributions. The result holds for all sample sizes. Asymptotic considerations appear in a later section.

The following version of the Neyman–Pearson lemma follows from the fact that $P(\mathbf{R} = \mathbf{r} \mid H_*, \boldsymbol{\tau}(\mathbf{x}))$ is constant for all $\mathbf{r}$ in $\{\mathbf{R} \mid \boldsymbol{\tau}(\mathbf{x})\}$ (Theorem 2.1).

Neyman–Pearson lemma for H_0. The most powerful size α conditional rank test (given $\boldsymbol{\tau}(\mathbf{x})$) for H_0 (or H_*) against some simple alternative H_a is given by

$$\begin{aligned}\Phi(\mathbf{r}) &= 1 \quad \text{if} \quad P(\mathbf{R} = \mathbf{r} \mid H_a, \boldsymbol{\tau}(\mathbf{x})) > k\,, \\ &= 0 \quad \text{if} \quad P(\mathbf{R} = \mathbf{r} \mid H_a, \boldsymbol{\tau}(\mathbf{x})) \leqq k\end{aligned}$$

where k is chosen so $E\{\Phi(\mathbf{R}) \mid \boldsymbol{\tau}(\mathbf{x})\}$ equals α under H_0 (or H_*).

For the following theorem define $a(t_i; F, \boldsymbol{\tau}(\mathbf{x}))$ for ranks t_i attainable in $\boldsymbol{\tau}(\mathbf{x})$, and for those functions $F(x; \theta)$ which are absolutely continuous with respect to $F(x; \theta_0)$, as

$$a(t_i; F_0, \tau(\mathbf{x})) = E\left\{\frac{\partial}{\partial\theta} f(X^{(t_i)}; \theta)|_{\theta=\theta_0} \mid \boldsymbol{\tau}(\mathbf{x}), H_0\right\}, \tag{6.1}$$

$$f(x; \theta) = \frac{dF(x; \theta)}{dF(x; \theta_0)}$$

where $X^{(j)}$ represents, as before, the jth order statistic in a random sample of size N from a population with cdf $F_0(x) = F(x; \theta_0)$.

THEOREM 6.1. *Let J be an open interval containing θ_0. If, for $\theta \in J$*

$$f(x; \theta) = \frac{dF(x; \theta)}{dF(x; \theta_0)} \quad \textit{exists}\,, \tag{6.2}$$

$$\frac{\partial}{\partial\theta} f(x; \theta)|_{\theta=\theta_0} = \lim_{\theta\to\theta_0} \frac{f(x; \theta) - f(x; \theta_0)}{\theta - \theta_0} \quad \textit{exists}\,, \tag{6.3}$$

and

$$f(x; \theta_0) = \lim_{\theta\to\theta_0} f(x; \theta) \quad \textit{exists}\,, \tag{6.4}$$

almost everywhere with respect to $F(x; \theta_0)$, and

$$\lim_{\theta\to\theta_0} \int_{-\infty}^{\infty} \left|\frac{\partial}{\partial\theta} f(x; \theta)\right| dF(x; \theta_0) = \int_{-\infty}^{\infty} \left|\frac{\partial}{\partial\theta} f(x; \theta)|_{\theta=\theta_0}\right| dF(x; \theta_0) < \infty \tag{6.5}$$

then the locally (small Δ) most powerful conditional (given $\boldsymbol{\tau}(\mathbf{x})$) rank test of H_0 against

$$H_a: F(x_1, \cdots, x_N) = \prod_{i=1}^N F(x_i: \theta_0 + \Delta c_i) \tag{6.6}$$

is given by the test with the critical region

$$\sum_{m=1}^N c_m a(R_m; F_0, \boldsymbol{\tau}(\mathbf{x})) > k \tag{6.7}$$

where k is chosen so the test will have size α for a given $\boldsymbol{\tau}(\mathbf{x})$.

PROOF. The proof resembles the proof in [10], page 71, except that the Radon–Nikodym derivative $f(x; \theta)$ is used instead of the usual density function, and the Stieltjes integral with element $\prod_{i=1}^N dF(x^{(r_i)}; \theta_0)$, over the region $x^{(t_1)} < \cdots < x^{(t_k)}$ is used instead of the usual Riemann integral. First the probabilities $P(\mathbf{R} = \mathbf{r} \mid H_a)$ are found to impose the same ordering, for small $(\theta - \theta_0)^2$, on the points $\{\mathbf{r}\}$ as that given by the statistic (6.7). This fact is then shown to hold when $\boldsymbol{\tau}(\mathbf{x})$ is given. The details are omitted.

7. The locally most powerful conditional rank test of H_1. A conditional rank test for H_1 is defined as a test based on $\mathbf{R}^+$ and sign $\mathbf{X}$, given $\boldsymbol{\tau}(|\mathbf{x}|)$. The following form of the Neyman–Pearson lemma follows immediately from the usual form of the Neyman–Pearson lemma, and the fact that $P(\mathbf{R}^+ = \mathbf{r}^+, \text{sign } \mathbf{X} = \mathbf{v} \mid \boldsymbol{\tau}(|\mathbf{x}|))$ is constant over all distributions in H_1 as given in (3.2).

Neyman–Pearson lemma for H_1. The most powerful size α conditional rank test for H_1 against some simple alternative H_a is given by

$$\begin{aligned}\Phi(\mathbf{r}^+, \mathbf{v}) &= 1 \quad \text{if} \quad P(\mathbf{R}^+ = \mathbf{r}^+, \text{sign } \mathbf{X} = \mathbf{v} \mid H_a, \boldsymbol{\tau}(|x|)) > k\,,\\ &= 0 \quad \text{if} \quad P(\mathbf{R}^+ = \mathbf{r}^+, \text{sign } \mathbf{X} = \mathbf{v} \mid H_a, \boldsymbol{\tau}(|x|)) \leqq k\end{aligned}$$

where k is chosen so $E\{\Phi(\mathbf{R}^+, \text{sign } \mathbf{X}) \mid \boldsymbol{\tau}(|\mathbf{x}|)\}$ equals α under H_1.
The following theorem shows the locally most powerful conditional rank test to be a linear rank test, under some regularity conditions:

THEOREM 7.1. *Under conditions* (6.2) *through* (6.5), *and*

$$\frac{\partial}{\partial\theta} f(-x;\theta)|_{\theta=\theta_0} = -\frac{\partial}{\partial\theta} f(x;\theta)|_{\theta=\theta_0} \tag{7.1}$$

the locally (small Δ) most powerful conditional rank test of H_1 against

$$H_a\colon F(x_1, \cdots, x_N) = \textstyle\sum_{i=1}^N F(x_i; \theta_0 + \Delta)$$

is given by the test with the critical region

$$\textstyle\sum_{m=1}^N \text{sign } X_m \cdot a_N^{+}(R_m^{+}; F_0, \boldsymbol{\tau}(|\mathbf{x}|)) > k \tag{7.2}$$

where k is chosen so the test will have size α for given $\boldsymbol{\tau}(|\mathbf{x}|)$, and where

$$a_N^{+}(t_m; F_0, \boldsymbol{\tau}(|\mathbf{x}|)) = E\left\{\frac{\partial}{\partial\theta} f(|X|^{(t_m)};\theta)|_{\theta=\theta_0} \mid \boldsymbol{\tau}(|\mathbf{x}|), H_1\right\}. \tag{7.3}$$

Note. Condition (7.1) ensures that the symmetry of H_1 is disturbed under H_a. The proof parallels that of Theorem 6.1, in conjunction with the proof on page 74 of [10].

8. The asymptotic distribution of S_0 under contiguous alternatives. The theorems of this section parallel the results in Chapter 6 of [10] for continuous random variables and in part Section 4 of [14] for discrete random variables. The principal difference is that the treatment involves no assumptions regarding the discrete or continuous nature of the random variables, although some regularity assumptions are made with respect to the parameter so that contiguous alternatives may be considered.

Consider a distribution function $F(x;\theta)$ with a parameter θ. As before let $f(x;\theta)$ represent the Radon–Nikodym derivative of $F(x;\theta)$ with respect to $F(x;\theta_0)$ and assume this derivative exists. Define the generalization of Fisher's information to be

$$I(F,\theta) = \int_{-\infty}^{\infty}\left[\frac{(\partial/\partial\theta)f(x;\theta)}{f(x;\theta)}\right]^2 dF(x;\theta)\,. \tag{8.1}$$

The distribution function of $Y = F(X; \theta)$ is denoted by $G(u; \theta)$, where $F(x; \theta)$ is the distribution function of X. Later we will find the following notation convenient:

$$\phi(u, F, \theta_0) = \frac{(\partial/\partial\theta) f(F^{-1}(u; \theta); \theta)|_{\theta=\theta_0}}{f(F^{-1}(u; \theta_0); \theta_0)} . \tag{8.2}$$

Although by definition $f(x; \theta_0) = 1$ we will often carry $f(x; \theta_0)$ through a development for clarity.

Denote the distribution function of the X's under H_0 by $F(x; \theta_0)$ and consider the alternative

H_{aN}: $X_1, \cdots, X_N$ are independent and X_i is distributed according to $F(x; \theta_i)$.

That is, the distribution functions of the X's differ in form through a single parameter. The asymptotic distribution of S_c is found under the local condition

$$\max_{1\leq i\leq N} (\theta_i - \theta_0)^2 \to 0 \tag{8.3}$$

and under the "non-triviality" condition

$$\lim_{N\to\infty} I(F, \theta_0) \textstyle\sum_{i=1}^N (\theta_i - \theta_0)^2 = b^2 \qquad \text{for} \quad 0 < b^2 < \infty \tag{8.4}$$

where Fisher's information is assumed to satisfy

$$0 < \lim_{\theta=\theta_0} I(F, \theta) = I(F, \theta_0) < \infty . \tag{8.5}$$

In addition, the regularity conditions (6.2), (6.3) and (6.4) are necessary. The double subscripts implied by the conditions (8.3) and (8.4) in order to properly approach a limit will be omitted for the sake of simplicity. Let

$$L_\theta = \textstyle\prod_{i=1}^N \dfrac{f(X_i; \theta_i)}{f(X_i; \theta_0)} \tag{8.6}$$

be the likelihood ratio and consider the statistics

$$W_\theta = 2 \textstyle\sum_{i=1}^N \left\{\left[\dfrac{f(X_i; \theta_i)}{f(X_i; \theta_0)}\right]^{\frac{1}{2}} - 1\right\} \tag{8.7}$$

and

$$T_\theta = \textstyle\sum_{i=1}^N (\theta_i - \theta_0)\phi(Y_i, F, \theta_0) , \tag{8.8}$$

THEOREM 8.1. *Assume that for θ in some open interval about θ_0,*

$$f(x; \theta) = \frac{dF(x; \theta)}{dF(x; \theta_0)} \quad \textit{exists} , \tag{8.9}$$

$$\frac{\partial}{\partial\theta} f(x; \theta)|_{\theta=\theta_0} = \lim_{\theta\to\theta_0} \frac{f(x; \theta) - f(x; \theta_0)}{\theta - \theta_0} \quad \textit{exists} , \tag{8.10}$$

and

$$f(x; \theta_0) = \lim_{\theta\to\theta_0} f(x; \theta) \quad \textit{exists} , \tag{8.11}$$

almost everywhere with respect to $F(x; \theta_0)$. Then (8.3), (8.4) *and* (8.5) *imply $T_\theta \to N(0, b^2)$ under H_0.*

The proof involves a direct application of Theorem V.1.2 of [10] and is omitted.

THEOREM 8.2. *The conditions of Theorem* 8.1 *imply*

$$\log L_0 - T_0 + \tfrac{1}{2}b^2 \to_P 0 \tag{8.12}$$

and

$$\log L_0 \to N(-\tfrac{1}{2}b^2, b^2) \tag{8.13}$$

under H_0.

PROOF. It is not difficult to show

$$E\{W_0\} \to -\tfrac{1}{4}b^2 \tag{8.14}$$

and

$$\operatorname{Var}\{W_0 - T_0\} \to 0 \tag{8.15}$$

in a manner similar to the proofs of Lemmas VI.2.1.a and VI.2.1.b of [10]. Thus

$$E\{(W_0 - T_0 + \tfrac{1}{4}b^2)^2\} \to 0 \tag{8.16}$$

and $W_0 \to N(-b^2/4, b^2)$ because of Theorem 8.1. LeCam's second lemma ([10], page 205) implies (8.12), and (8.13) immediately follows.

The following theorem is the main result of this section, as it presents the distribution of S_c under contiguous alternatives and is thus analogous to Theorem VI.2.4 of [10]. For purposes of this section it is more convenient to use

$$S_c' = S_c - E\{S_c \mid T_N\} = \textstyle\sum_{i=1}^N (c_i - \bar{c})a_N(R_i; T_N) \tag{8.17}$$

whose limiting distribution under H_0 (see Theorem 4.1) is $N(0, \sigma_c^2)$ where σ_c^2 is given by (4.13).

THEOREM 8.3. *Let* $\phi(u)$ *be a non-constant square integrable function, on* $0 \leqq u \leqq 1$, *and assume*

$$\textstyle\int_0^1 (a_N(N \cdot T_N^{-1}(u); T_N) - \phi(u))^2 \, du \to_P 0 \tag{8.18}$$

holds under H_0. *Then if*

$$\textstyle\sum_{i=1}^N (c_i - \bar{c})^2 / \max_{1 \leqq i \leqq N} (c_i - \bar{c})^2 \to \infty , \tag{8.19}$$

holds, the conditions of Theorem 8.1 *imply that* S_c' *is asymptotically* $N(\mu_{0c}, \sigma_c^2)$ *under* H_a, *where*

$$\mu_{0c} = \textstyle\sum_{i=1}^N (c_i - \bar{c})(\theta_i - \theta_0) \int_0^1 \phi(u)\phi(u, F, \theta_0)\, du \tag{8.20}$$

and

$$\sigma_c^2 = \textstyle\sum_{i=1}^N (c_i - \bar{c})^2 \int_0^1 (\phi(u) - \bar{\phi})^2 \, du . \tag{8.21}$$

PROOF. It was shown in the proof of Theorem 4.1, by (4.16) and (4.18), that S_c' is asymptotically equivalent under H_0 to T_c, defined by (4.12), and hence may be replaced by T_c in asymptotic considerations. This and (8.12) imply that the bivariate random variables $(S_c', \log L_0)$ and $(T_c, T_0 - b^2/2)$ converge in probability to each other. Under H_0 we have $T_c \to N(0, \sigma_c^2)$ and $T_\theta \to N(0, b^2)$. Note that T_θ can also be written as

$$T_\theta = \sum_{i=1}^N (\theta_i - \theta_0)\phi(U_i, F, \theta_0) \tag{8.22}$$

where U_i, defined by (4.11), is the same uniformly distributed random variable which appears in T_c. Thus the covariance of T_c and T_θ is

$$\operatorname{Cov}(T_c, T_\theta) = \sum_{i=1}^N (c_i - \bar{c})(\theta_i - \theta_0) \int_0^1 \phi(u)\phi(u, F, \theta_0)\, du \tag{8.23}$$

because $E\{T_\theta\} = 0$. The remainder of the proof that (T_c, T_θ) is asymptotically bivariate normal is the same as on page 218 of [10]. But this implies $(S_c', \log L_\theta)$ is asymptotically bivariate normal under H_0, and the parameters of the asymptotic distribution satisfy the conditions of LeCam's third lemma, page 208 of [10], which states that under H_a S_c' is asymptotically normal (μ_{0c}, σ_c^2).

In the case of several samples and Q statistics we have the following analogue to Theorem VI.3.1 of [10].

THEOREM 8.4. *Under the conditions of Theorem* 8.3, *but with condition* (8.19) *replaced by* $\min(n_1, \cdots, n_k) \to \infty$, *the statistic* $Q_{n_1,\cdots,n_k}$ *defined by* (4.33) *has asymptotically the noncentral chi-square distribution with* $k - 1$ *degrees of freedom and noncentrality parameter*

$$\delta = \frac{\sum_{j=1}^k ((\sum_{i \in s_j} (\theta_i - \bar{\theta}))^2/n_i)(\int_0^1 \phi(u)\phi(u, F, \theta_0)\, du)^2}{\int_0^1 (\phi(u) - \bar{\phi})^2\, du}. \tag{8.24}$$

PROOF. The proof is similar to the proof of Theorem VI.3.1 of [10], except that Theorem 8.3 is used instead of VI.2.4 from [10].

9. The asymptotic distribution of S_N^+ under contiguous alternatives. For the following development it will be convenient to use the function

$$\phi^+(u, F, \theta_0) = \frac{\partial}{\partial\theta} f(F^{-1}(\tfrac{1}{2} + \tfrac{1}{2}u); \theta)|_{\theta=\theta_0}. \tag{9.1}$$

It is interesting to note that

$$\phi^+(u, F, \theta_0) = \phi(\tfrac{1}{2} + \tfrac{1}{2}u, F, \theta_0) \tag{9.2}$$

where the latter function is defined by (8.2).

Let $F(x; \theta)$ be a symmetric function for $\theta = \theta_0$ (when H_1 is true) and define the likelihood function

$$L_\Delta = \prod_{i=1}^N \frac{f(X_i; \theta_0 + \Delta)}{f(X_i; \theta_0)}. \tag{9.3}$$

Define $I(F, \theta)$ as in (8.1), and assume

$$\Delta \to 0 \tag{9.4}$$

$$\lim_{N\to\infty} I(F, \theta_0) \cdot N\Delta^2 = b^2 \qquad \text{for} \quad 0 < b^2 < \infty \tag{9.5}$$

and

$$0 < \lim_{\theta\to\theta_0} I(F, \theta) = I(F, \theta_0) < \infty. \tag{9.6}$$

Then the counterpart to T_θ in (8.8) is

$$T_\Delta = \sum_{i=1}^N \Delta\phi(F(X_i), F, \theta_0) = \sum_{i=1}^N \Delta \frac{\partial}{\partial\theta} f(X_i; \theta)|_{\theta=\theta_0} \tag{9.7}$$

which can also be written as

$$T_\Delta = \sum_{i=1}^N \Delta \operatorname{sign} X_i \frac{\partial}{\partial\theta} f(|X_i|; \theta)|_{\theta=\theta_0} \tag{9.8}$$

$$= \sum_{i=1}^N \Delta \operatorname{sign} X_i \phi^+(U_i^+, F, \theta_0) \tag{9.9}$$

where U_i^+ is defined in Section 5, if (7.1) is true.

THEOREM 9.1. *Let* $F(x; \theta)$ *satisfy* (6.2), (6.3), (6.4), (7.1) *and* (9.6). *If* (5.5) *holds under* H_1 *for some square integrable* (*on* (0, 1)) *function* $\phi^+(u)$ *that satisfies* (5.4), *then* (9.4) *and* (9.5) *imply that the sequence* S_N^+ *is asymptotically* $N(\mu_\Delta, \sigma^2)$ *under* H_{aN}, *where* σ^2 *is given by* (5.6) *and* μ_Δ *by*

$$\mu_\Delta = N\Delta \int_{F^+(0)}^1 \phi^+(u)\phi^+(u, F, \theta_0)\, du\,, \tag{9.10}$$

and the sequence S_N^+/σ_N (*see* (5.6)) *is asymptotically* $N(\mu_\Delta/\sigma, 1)$.

PROOF. The proof is analogous to the proof of Theorem 8.3 and is therefore omitted.

10. Asymptotic relative efficiency. When testing H_0, if there is convergence

$$\frac{\sum_{i=1}^N (c_i - \bar{c})(\theta_i - \theta_0)}{(\sum_{i=1}^N (c_i - \bar{c})^2 \sum_{i=1}^N (\theta_i - \theta_0)^2)^{\frac{1}{2}}} \to \rho_2 \tag{10.1}$$

then the *asymptotic efficiency* of the test using S_c is defined in [10] as

$$e = \rho_1^2 \rho_2^2 \tag{10.2}$$

where ρ_1 is given by

$$\rho_1 = \frac{\int_0^1 \phi(u)\phi(u, F, \theta_0)\, du}{(\int_0^1 (\phi(u) - \bar{\phi})^2\, du \int_0^1 \phi^2(u, F, \theta_0)\, du)^{\frac{1}{2}}}\,. \tag{10.3}$$

This notation enables the asymptotic distribution of S_c' under the conditions of Theorem 8.3 to be stated simply as $N(\rho_1\rho_2 b\sigma_c, \sigma_c^2)$, and the asymptotic distribution of $(S_c - E\{S_c \mid T_N\})/(\operatorname{Var}\{S_c \mid T_N\})^{\frac{1}{2}}$ under the same conditions to be $N(\rho_1\rho_2 b, 1)$.

The ratio of the two asymptotic efficiencies of two comparable tests is called their *asymptotic relative efficiency* (ARE). In the usual case of interest the constants c_i are the same for both tests but the $\phi(u)$ functions differ. Then the ARE of the test using $\phi_1(u)$, say, relative to the test using $\phi_2(u)$ is

$$\text{ARE} = \frac{(\int_0^1 \phi_1(u)\phi(u, F, \theta_0)\, du)^2 \int_0^1 (\phi_2(u) - \bar{\phi}_2)^2\, du}{(\int_0^1 \phi_2(u)\phi(u, F, \theta_0)\, du)^2 \int_0^1 (\phi_1(u) - \bar{\phi}_1)^2\, du}\,. \tag{10.4}$$

Thus the asymptotic efficiency is the ARE relative to the most powerful test, under the given conditions, and the most powerful test is one which uses $c_i = \theta_i$ and $\phi(u) = \phi(u, F, \theta_0)$, or some linear functions thereof.

We may now draw some immediate conclusions regarding the asymptotic efficiency of S_c' for the various methods of scoring ties discussed in Section 4.

Average score method. If the conditions of Theorem 4.2 are met for the average scores defined by (4.20), then under the conditions of Theorem 8.3 the distribution of $(S_c - E\{S_c \mid T_N\})/(\text{Var}\{S_c \mid T_N\})^{\frac{1}{2}}$ is asymptotically normal with mean $\rho_1\rho_2 b$ and variance 1, where ρ_1 is given by

$$(10.5) \qquad \rho_1 = \frac{\int_0^1 \phi_a(u)\phi(u, F, \theta_0)\, du}{(\int_0^1 (\phi_a(u) - \bar{\phi})^2\, du\, I(F, \theta_0))^{\frac{1}{2}}}$$

for $\phi_a(u)$ defined by (4.21). In this case our result is a generalization of Theorem 4.2 of [14].

Midrank method. When the midrank scores of (4.28) satisfy the conditions of Theorem 4.3, then the asymptotic distribution of $(S_c - E\{S_c \mid T_N\})/(\text{Var}\{S_c \mid T_N\})^{\frac{1}{2}}$ under the conditions of Theorem 8.3 is $N(\rho_1\rho_2 b, 1)$ where

$$(10.6) \qquad \rho_1 = \frac{\int_0^1 \phi_m(u)\phi(u, F, \theta_0)\, du}{(\int_0^1 (\phi_m(u) - \bar{\phi}_m)^2\, du\, I(F, \theta_0))^{\frac{1}{2}}}$$

for $\phi_m(u)$ defined in the text preceding Theorem 4.3.

Randomized ranks. In the case of randomized ranks the asymptotic distribution of $(S_c - E\{S_c \mid T_N\})/(\text{Var}\{S_c \mid T_N\})^{\frac{1}{2}}$ under the conditions of Theorems 4.4 and 8.3 is $N(\rho_1\rho_2 b, 1)$ where

$$(10.7) \qquad \rho_1 = \frac{\int_0^1 \phi(u)\phi(u, F, \theta_0)\, du}{(\int_0^1 (\phi(u) - \bar{\phi})^2\, du\, I(F, \theta_0))^{\frac{1}{2}}}\, .$$

Note that the numerators of both ρ_1's defined by (10.5) and (10.7) are identical, since $\phi(u, F, \theta_0)$ is constant over the same intervals in which $\phi(u)$ is averaged to give $\phi_a(u)$. Therefore the ARE of an average scores test relative to a randomized ranks test is

$$(10.8) \qquad \text{ARE} = \frac{\int_0^1 (\phi(u) - \bar{\phi})^2\, du}{\int_0^1 (\phi_a(u) - \bar{\phi})^2\, du} \geqq 1.0$$

with equality only if $\phi(u)$ is constant over the same intervals in which $G(u)$ is constant, so that $\phi(u) \equiv \phi_a(u)$. Note also that if $\phi(u)$ is constant over the same intervals in which $G(u)$ is constant, such as when $\phi(u) = \phi(u, F, \theta_0)$ in the most powerful test, then the average scores method, the midrank method, and the method of randomized ranks have identical asymptotic efficiencies, because then $\phi_a(u) \equiv \phi_m(u) \equiv \phi(u)$.

In the usual multi-sample case of interest, the alternative hypothesis specifies that θ_i is constant within the partitions s_j; that is, we have k populations;

$$(10.9) \qquad H_a\colon F(x_1, \cdots, x_N) = \prod_{j=1}^k \prod_{i \in s_j} F(x_i; \theta_j)\, .$$

Then the asymptotic efficiency of the Q test, as defined on page 271 of [10], is

$$(10.10) \qquad e = \rho_1^2 = \frac{(\int_0^1 \phi(u)\phi(u, F, \theta_0)\, du)^2}{I(F, \theta_0) \int_0^1 (\phi(u) - \bar{\phi})^2\, du}$$

which equals 1.0 if $\phi(u) = \phi(u, F, \theta_0)$. Note that $\phi(u)$ in (8.24) and (10.10) is replaced by $\phi_a(u)$ or $\phi_m(u)$ in case average scores or midrank scores are used.

For tests of symmetry involving S_N^+ the asymptotic efficiency becomes

$$e = \frac{[\int_{F^+(0)}^1 \phi^+(u)\phi^+(u, F, \theta_0)\,du]^2}{\int_{F^+(0)}^1 [\phi^+(u)]^2\,du \int_{F^+(0)}^1 [\phi^+(u, F, \theta_0)]^2\,du} \tag{10.11}$$

which corrects the equation on page 277 of [10]. Thus the most efficient test is one whose scores converge, in the sense of equation (5.5), to $\phi^+(u, F, \theta_0)$, defined by (9.1). Comparisons among the various methods for breaking ties may be made here also, with the same results as before.

11. Applications of the tests for randomness. The function $\phi(Y_i, F, \theta_0)$ is important in the discussion of most powerful tests for randomness. Theorem 6.1 showed that the locally most powerful test of H_0 is based on the statistic

$$S_c^\phi = \sum_{i=1}^N c_i E\{\phi(Y^{(R_i)}, F, \theta_0) \mid T_N, H_0\} \tag{11.1}$$

while Theorem 8.3 implies that the asymptotically most powerful test is based on the same statistic if the correlation between c_i and θ_i is unity. Actually Theorem 8.3 implies that any linear rank statistic $S_c = \sum_{i=1}^N c_i a(R_i; T_N)$ which uses scores that converge to $\phi(u, F, \theta_0)$ in the sense of Theorem 8.3 is asymptotically equivalent to (11.1). One such statistic is

$$S_c = \sum_{i=1}^N c_i \phi_a\left(\frac{R_i}{N+1}, F, \theta_0 \mid T_N\right) \tag{11.2}$$

which uses the average of the scores $\phi(i/(N+1), F, \theta_0)$ in the sense of (4.20).

In many cases $\phi(u, F, \theta_0)$ is a linear function of $F^{-1}(u; \theta_0)$. In particular

$$\phi(u, F, \theta_0) = \frac{F^{-1}(u; \theta_0) - E\{X_i\}}{\operatorname{Var}\{X_i\}} \tag{11.3}$$

holds for the following cases.

(a) $F(x; \theta)$ is Poisson, with $\theta = E(X) = \lambda$;

(b) $F(x; \theta)$ is binomial, with $\theta = E(X) = np$, constant n;

(c) $F(x; \theta_0)$ is uniform on (0, 1) while $F(x; \theta)$ considers the alternative where the density equals $1 + (2x - 1)\delta$ for x in (0, 1) and zero elsewhere, for a parameter δ, and $\theta = E(X) = \frac{1}{2} + \delta/6$, $(\theta_0 = \frac{1}{2})$;

(d) $F(x; \theta)$ is geometric with probabilities $(1-p)p^k$ for $k = 0, 1, 2, \cdots$, and with $\theta = E(X) = p/(1-p)$;

(e) $F(x; \theta)$ is exponential, with $\theta = E(X) = 1/\lambda$;

(f) $F(x; \theta)$ is normal, with $\theta = E(X) = \mu$ and σ constant.

In the above cases it is equivalent to use $F^{-1}(u; \theta_0)$ for $\phi(u)$ to obtain an asymptotically most powerful rank test. That is, because of Corollary V.1.6 of [10], the scores

$$a_N(i) = \text{the } \frac{i}{N+1} \text{ quantile of } F(x; \theta_0) \tag{11.4}$$

may be used in the average scores method, the mid-rank method, or the randomized ranks method to obtain an asymptotically most powerful rank test for the cases listed. The above list is not intended to be exhaustive.

The flexibility now allowed by rank tests may be illustrated by the following example of a mixture of discrete and continuous distributions. Consider the distribution function of X,

$$(11.5) \qquad \begin{aligned} F(x) &= 0 \qquad \text{if} \quad x < 0\,, \\ &= 1 - p + p \textstyle\int_0^x g(t)\,dt \qquad \text{if} \quad x \geqq 0 \end{aligned}$$

where $g(t)$ is some density function on $(0, \infty)$ with mean b, say. Such a distribution function $F(x)$ could be used to describe monthly volumes of water carried by streams which are dry during long periods, or individual incomes in a population where not everyone has an income. If $\theta = p$, $\theta_0 = p_0$, and b is constant, then

$$(11.6) \qquad \begin{aligned} \phi(u, F, E(X)) &= \frac{-1}{b(1-p_0)} \qquad \text{if} \quad 0 < u \leqq 1 - p_0\,, \\ &= \frac{1}{bp_0} \qquad \text{if} \quad 1 - p_0 < u \leqq 1\,. \end{aligned}$$

This indicates that the asymptotically most powerful linear rank test of randomness uses scores given by (11.6) in a statistic such as S_c defined by (11.2) or, equivalently, simply

$$(11.7) \qquad S_c = \textstyle\sum_{i=1}^N \operatorname{sign}(X_i)$$

where sign (X_i) equals 0 or 1 depending on whether X_i equals 0 or is positive.

Usually a most powerful test is described for alternatives which include a class of distributions, such as the class of Poisson distributions with unspecified parameter λ. In such a case it is permissible to estimate the unspecified parameter or parameters from $\mathbf{X}^{(\cdot)}$, the combined ordered sample, because $\mathbf{X}^{(\cdot)}$ is, for a given vector of ties, independent of the ranks $\mathbf{R}$ by virtue of Theorem 2.1. If the estimator used is consistent, then asymptotically it is equivalent to the true value of the parameter and the asymptotic efficiency of the test is not impaired. In fact the entire function may be estimated from $X^{(\cdot)}$, perhaps in a manner similar to that in Theorem VII.1.5 of [10], to obtain a test that is asymptotically efficient against all alternatives. The details of such a test in the noncontinuous case have not been worked out.

In closing we should add a postscript to the remark that the asymptotic efficiencies of the average scores method, the midrank method and the randomized ranks method all equal 1.0 when the most powerful test is being used. If we are using the wrong scores because of ignorance concerning the true distribution function $F(x; \theta_0)$, or if we are deliberately using nonoptimum scores for convenience, then the asymptotic efficiency of the randomized ranks method is likely to be inferior to that of the average scores method. That is, the latter test will be more powerful whenever the function being used is not constant over all constant intervals of $G(u; \theta_0)$, such as when using the Mann–Whitney test in the presence of ties. It is likely that simple conditions exist under which the midrank method is superior to the randomized ranks method, but such conditions have not been established. A comparison between the midranks method and the average scores method has been made in [7], in which comparisons are made between the results of this paper and Table 1 of [4].

REFERENCES

[1] ADICHIE, J. N. (1967). Asymptotic efficiency of a class of nonparametric tests for regression parameters. *Ann. Math. Statist.* **38** 884–893.

[2] ANDREWS, F. C. and TRUAX, D. R. (1964). Locally most powerful rank tests for several sample problems. *Metrika* **8** 16–24.

[3] BERAN, R. J. (1970). Linear rank statistics under alternatives indexed by a vector parameter. *Ann. Math. Statist.* **41** 1896–1905.

[4] BÜHLER, W. J. (1967). The treatment of ties in the Wilcoxon test. *Ann. Math. Statist.* **38** 519–522.

[5] CHANDA, K. C. (1963). On the efficiency of two-sample Mann-Whitney test for discrete populations. *Ann. Math. Statist.* **34** 612–617.

[6] CHERNOFF, H. and SAVAGE, I. R. (1958). Asymptotic normality and efficiency of certain nonparametric test statistics. *Ann. Math. Statist.* **29** 972–994.

[7] CONOVER, W. J. and KEMP, K. E. (1971). How to use rank tests in the presence of ties. Technical Report No. 21, Kansas State Univ.

[8] HÁJEK, J. (1961). Some extensions of the Wald-Wolfowitz-Noether theorem. *Ann. Math. Statist.* **32** 506–523.

[9] HÁJEK, J. (1962). Asymptotically most powerful rank order tests. *Ann. Math. Statist.* **33** 1124–1147.

[10] HÁJEK, J. and ŠIDÁK, Z. (1967). *Theory of Rank Tests*. Academic Press, New York.

[11] LECAM, L. (1960). Locally asymptotically normal families of distributions. *Univ. California Publ. Statist.* **3** 37–98.

[12] MATTHES, T. K. and TRUAX, D. R. (1965). Optimal invariant rank tests for the k sample problem. *Ann. Math. Statist.* **36** 1207–1222.

[13] PUTTER, J. (1955). The treatment of ties in some nonparametric tests. *Ann. Math. Statist.* **26** 368–386.

[14] VORLÍČKOVÁ D. (1970). Asymptotic properties of rank tests under discrete distributions. *Z. Wahrscheinlichkeitstheorie und verw. Gebiete* **14** 275–289.

[15] VORLÍČKOVÁ, D. (1972). Asymptotic properties of rank tests of symmetry under discrete distributions. *Ann. Math. Statist.* **43** 2013–2018.

DEPARTMENT OF MATHEMATICS
TEXAS TECH UNIVERSITY
LUBBOCK, TEXAS 79409